Optimal Stopping Problems Under Partial Information with Applications in Finance

Oluwafunmilola Oluwatobiloba Olusanya (funmilola@aims.ac.za)
African Institute for Mathematical Sciences (AIMS)

Supervised by: Dr. Dennis Ikpe
University of South Africa (UNISA), South Africa

Co-supervised by: Mr Aidoo Edward
Barclay's Africa Group, South Africa

18 May 2017

Submitted in partial fulfillment of a structured masters degree at AIMS South Africa

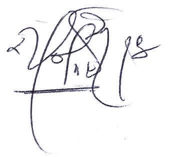
Abstract

We study the optimal stopping problem under partial information with square-integrable Lévy process. In order to solve this, we convert it to a full information optimal stopping problem by constructing a Kalman-Bucy filter of the unobservable process. This filter is obtained as estimates of the mean and conditional variance. We solve the resulting optimal stopping problem under full information by using Snell envelope theory and the dynamic programming principle. We then demonstrate an application in mathematical finance for an American type option with stochastic convenience yield.

Keywords: Optimal stopping problem, Square-integrable, Kalman-Bucy filter, Dynamic programming principle.

Declaration

I, the undersigned, hereby declare that the work contained in this research project is my original work, and that any work done by others or by myself previously has been acknowledged and referenced accordingly.



Olusanya Oluwafunmilola Oluwatobiloba, 18 May 2017

Contents

1. Introduction

This project focuses on optimal stopping problems under partial information with observable and un-observable equations driven by square integrable Lévy processes. Optimal stopping problems are a very salient and esteemed class of optimal control problems in finance and they involve making a decision on when best to stop a given process. Our aim in addressing these type of problems is two fold. The first is to obtain an optimal stopping criterion for stopping a functional of an unobservable stochastic process in order to minimize risk and maximize the value of the expected payoff. The second is to be able to apply these problems to real world situations, particularly in mathematical finance.

The concept and theory of optimal stopping problems dates back to the 1970's when Fischer Black and Myron Scholes came up with a formula for estimating stock option prices (Hill, 2009). Ever since, the study of these type of problems has gained popularity among researchers and financial analysts. Their applications include but are not limited to; selling of financial derivatives, pricing of options, optimal stock selling and quality control and reliability. These can be found in (Vannestål, 2011), (Rishel and Helmes, 2006), (Wang, 2017) and (Jensen and Hsu, 1993).

Optimal stopping problems under partial information are usually difficult to solve since we do not have full information and we would therefore require some inference from a known observable process adapted to a given filtration while at the same time working towards making an optimal decision.

The theory of optimal stopping under partial information has been studied in different numerical contexts. Zhou (2013) put forward a Near value iteration method for the solution of continuous-state optimal stopping problems that defy exact solutions and pose some difficulties when handled numerically. The aim of his method was to derive a value function approximation (and it's iteration) close enough to the true value function while at the same time achieving computational savings. He later worked with Fan Ye and together, they devised a numerical approach for tackling the optimal stopping problem of partially observable Markov processes. Their approach tagged; filtering-based duality approach, relied on some martingale duality formulation and Kalman filtering technique aimed at obtaining an asymptotic upper bound on the value function (see (Ye and Zhou, 2013)). Some other important literatures are; (Ludkovski, 2009) and (Jensen and Hsu, 1993).

In this project, we propose to solve the partial information-based optimal stopping problem of square integrable observable and non-observable Lévy processes by first converting them to a full information problem through the utilization of the Kalman-Bucy filtering technique. While this technique has focused on the construction of a filter [see (Applebaum and Blackwood, 2015) and (Wang and Wu, 2008)], there is no existing literature to our knowledge on the specific observable and un-observable process as well as the optimal stopping application considered in this project. We then go on to solve the resulting full information problem by means of the dynamic programming principle and also demonstrate an application in mathematical finance.

The key mathematical contribution of this project lies in the derivation of a finite dimensional filter as well as the proof that the Snell envelope satisfies the dynamic programming principle.

While the optimal stopping problem presented in this project is mathematically interesting in its own right, we emphasize that there is a room for application to American option under stochastic convenience yield.

The rest of this project is organised as follows; Later in Chapter one, we give a brief introduction of square integrable Lévy processes and a formal problem statement of our work. We also state some important applications of optimal stopping problems. In Chapter two, we discuss the Kalman-Bucy filter in the context of square integrable Lévy processes and introduce a filtering model coupled with an innovation process through which a filtering distribution is obtained, and hence the transformation of the partial information optimal stopping problem to a corresponding full information problem. We also discuss the solution of the latter by means of the dynamic programming principle. In Chapter three, we consider an application of the optimal stopping problem and finally in Chapter four, we give a conclusion of the project .

1.1 Square Integrable Lévy Processes

Before discussing Square integrable Lévy processes, we shall give a formal definition of Lévy processes in the spirit of (Papapantoleon, 2008).

1.1.1 Definition. (Lévy Process) Let $(\Omega, \mathcal{F}, \mathbb{F}, P)$ be a filtered probability space with filtration $\mathbb{F} := (\mathcal{F}_t)_{t\in[0,T]}$ and let $T \in [0, \infty]$ be a given time horizon. Then we define a Lévy process to be a real-valued stochastic process $L = (L_t)_{0\leq t\leq T}$ that satisfies the following conditions;

1. $L_0 = 0$ almost surely.
2. L has independent increments i.e for $t_1,\ t_2, \ldots,\ t_n \in T$ with $t_1 < t_2 < \cdots < t_n$, we have that $L_{t_1},\ L_{t_2} - L_{t_1},\ \ \ldots, L_{t_n} - L_{t_{n-1}}$ are independent.
3. L has stationary increments, i.e for $s, t \in T$ with $s \leq t$, both L_{t-s} and the increment $L_t - L_s$ have the same distribution.
4. L is stochastically continuous, i.e for every $0 \leq t \leq T$ and $\epsilon > 0$, we have that; $\lim_{s\to t} P(|L_t - L_s| > \epsilon) = 0$.

All Lévy processes are usually right continuous with left limits (i.e Càdlàg) and they have lots of applications in finance such as in Lévy-based stochastic volatility models and models for stock price among others. Some examples of Lévy processes are the Gaussian process (i.e Brownian motion with drift), Poisson process and Compound Poisson process. In this project, we are concerned with square integrable Lévy processes.

1.1.2 Definition. Let $L \in M(\mathbb{R}^d)$ be a Lévy process. Then L is said to be square integrable if, for every test function $\phi \in N(\mathbb{R}^d)$, the real valued process $(\langle L(t), \phi\rangle, t \geq 0)$ is square integrable (see (Peszat and Zabczyk, 2007)).

We can express a square integrable Lévy process L in the form:

$$L_t = \mu_t + \sigma_t W_t + \int_0^t \int_{\mathbb{R}} x\tilde{N}(ds, dx).$$

where $\mu \in \mathbb{R}$ is the drift term, $\sigma_t \in Z^+$ is the stochastic volatility, W is a wiener process and "$\tilde{N}$ is the compensated random measure of a Poisson process associated with a Poisson random measure N of L" (see (Rémillard and Renaud, 2011)).

1.2 Problem Statement

Let $(\Omega, \mathcal{F}, P)$ be a probability space that supports a continuous Lévy-driven random process with a filtration $\mathbb{F} := \{\mathcal{F}_t\}_{t\geq 0}$ that satisfies the usual conditions. We shall consider an observable process and an unobservable process $\{(X_t, Y_t),\ t = 0, 1, \dots, T\}$ with the following dynamics;

$$dX_t = \lambda_t X_t \ \mathrm{dt} + \eta_t \mathrm{dL}_t^2 + \mathrm{dL}_t^1, \tag{1.2.1}$$

$$dY_t = \sigma_t X_t \ \mathrm{dt} + \mathrm{dL}_t^2. \tag{1.2.2}$$

- λ_t is a locally-bounded left continuous real-valued function,
- $(L_t^1)_{0\leq t\leq T}$ and $(L_t^2)_{0\leq t\leq T}$ are two independent square integrable Lévy processes with finite variance and are usually regarded to as the noise terms,
- $\{X_t\}$ represents the unobservable process,
- σ_t and η_t are locally bounded left continuous functions taking values in $\mathbb{R}$. Where the deterministic and bounded function σ_t gives a measure of how the actual value of the unobservable process is revealed over time,
- $\{Y_t\}$ is the observable process and,
- $\{(X_t,\ Y_t\ )\}$ is a bivariate process adapted to the filtration $\{\mathcal{F} := \sigma(X_i,\ Y_i);\ i = 0, \dots, t\}$.

The dynamics in (1.2.1) is referred to as the state equation, whereas, the dynamics in (1.2.2) is the observation equation. We shall denote the filtration generated by Y up until time t as $\{{\mathcal{F}_t}^Y := \sigma(Y_0, \dots, Y_s),\ s \leq t\}$ and the filtration generated by X as ${\mathcal{F}_t}^X$ respectively. In a similar manner, we shall take $\mathcal{T}^Y$ to be the set of ${\mathcal{F}_t}^Y$ stopping times that take values in $K = \{1, \dots, T\}$.

Suppose the initial value of Y_0 is known and we make the assumption that the initial unobservable process X_0 follows a distribution say π_0 which is known and formulated from past data (that includes Y_0), then we can consider the finite-horizon partially observable optimal stopping problem given as

$$S_0(\pi_0,\ y_0) = \sup_{\tau\in\mathcal{T}^Y} \mathbb{E}\ [\ F(\tau, X_\tau, Y_\tau)\ |\ X_0 \sim \pi_0,\ Y_0 = y_0], \tag{1.2.3}$$

where the payoff function F is a function of the random variables τ, the process X and the observed process Y. The right hand side of (1.2.3) is usually difficult to determine because we can't observe X_τ. This implies that the stopping time $\tau \in \mathcal{T}^Y$ is not adapted to ${\mathcal{F}_t}^{(X,Y)}$. Thus, it would be difficult to perform full optimal stopping on (1.2.1). This difficulty can be overcome by harnessing the availability of some noisy information process $\{Y_t\}_{t\in[0,T]}$ that contains some partial information of the unobservable process and is adapted to the filtration $\{{\mathcal{F}_t}^Y\}$. Thus, we have the problem of optimal stopping of (1.2.1) given that we observe (1.2.2) (i.e optimal stopping problem under partial information).

Suppose in (1.2.3) that τ is also adapted to ${\mathcal{F}_t}^{(X,Y)}$, then we would have the problem of optimal stopping under full information which can then be solved directly using dynamic programming principle or other methods but this is not the case. Thus, the objective of this project is to achieve the following;

1. Convert the optimal stopping problem under partial information in (1.2.3) to an optimal stopping problem under full information by carrying out some filtering on X given the observation Y.

2. Determine some optimal stopping criterion by using Snell envelope theory and the dynamic programming principle.

3. Consider an application in mathematical finance.

The first target can be achieved by constructing a filter on X_t which is a conditional expectation defined as

$$\Pi_t(Y_t) = \mathbb{E}(\ g(X_t) \mid \mathcal{F}_t{}^Y), \tag{1.2.4}$$

where the real-valued function, g, is any payoff function which is locally lipshitz[1] and has at most quadratic growth.

Equation (1.2.4) above is a filtering problem given that we have the observation process Y_t. For general Lévy processes, the filtering equation of this kind suffers from the "curse of dimensionality". It has been shown in (Ikpe and Becker, 2016), that (1.2.4) is finite dimensional for our observation process. Thus in this project, we would not consider the case of infinite-dimensionality. However, it will be interesting to extend our current result to the particular case of infinite second moment in the future .

Using Theorems 2.1 and 2.2 in (Ikpe and Becker, 2016), we will be able to reduce the filtering distribution Π_t to a two dimensional process. Theorem 2.1 gives the dynamics of the variance, whereas, theorem 2.2 gives the dynamics of the mean. Thus the filtering recursion can be re-written as

$$\Pi_t = \varphi\ (\Pi_{t-1},\ Y_{t-1},\ Y_t),\ t = 1, 2, \ldots, T.$$

We can now convert the optimal stopping problem under partial information in (1.2.3) to an equivalent optimal stopping problem under the fully observed state $(\Pi_t,\ Y_t)$ i.e.

$$S_0(\pi_0, y_0) = \sup_{\tau \in \mathcal{T}^Y} \mathbb{E}[\ F(\Pi_\tau, Y_\tau) \mid X_0 \sim \pi_0,\ Y_0 = y_0], \tag{1.2.5}$$

where;

$$F(\Pi_t, Y_t) :=\ \mathbb{E}[\ F(X_t,\ Y_t) \mid \mathcal{F}_t{}^Y]$$

By dynamic programming principle, we can then perform the backward iteration

$$S_t(\Pi_t,\ Y_t) = \max\left(F(\Pi_t, Y_t),\ C_t\ (\Pi_t,\ Y_t)\right),\ t = T, \ldots, 1, \tag{1.2.6}$$

where; $F(\Pi_t,\ Y_t)$ is an input in the problem and C_t is the continuation value at time t which would be explicitly obtained iteratively from the terminal point as

$$C_T(\Pi_T, Y_T) = F(\Pi_T, Y_T).$$

So that,

$$C_{t-1}(\Pi_{t-1}, Y_{t-1}) = \mathbb{E}[S_t(\Pi_t,\ Y_t) \mid \Pi_{t-1},\ Y_{t-1}],\ t = T-1, \ldots, 0.$$

Where S_t is computable from (1.2.6), $S_0 = C_0$ and the optimal stopping time is defined as;

$$T^* = \min\ \{t \in \mathcal{K} \mid F(\Pi_t,\ Y_t) \geq C_t(\Pi_t,\ Y_t)\}.$$

[1] A function $f : B \subset \mathbb{R}^n \to \mathbb{R}^m$ is locally lipshitz at $y_0 \in B$ if there exists constants $\delta > 0$ and $M > 0$ such that whenever $||x - x_0|| < \delta$, then $||f(x) - f(x_0)|| \leq M||x - x(0)||$.

Thus Π_t is a sufficient statistic for determining the conditional distribution $X_t|\mathcal{F}_t{}^Y$.

Some Applications of Optimal Stopping Problem in Finance

Below are two real world applications of Equations (1.2.1), (1.2.2) and (1.2.5) as seen in (Ludkovski, 2009).

1. Optimal Investment Under Partial Information.

The problem of investment timing usually arise in the theory of real options. For instance, a director may decide to embark on a new enterprise with worth (Y_t) which changes according to

$$dY_t = X_t\ dt + \sigma_Y\ dB_t.$$

Where (X_t) is the drift parameter and dB_t is an $\mathbb{R}$-valued Brownian motion with continuous density. The asset process Y_t follows the same dynamics in (1.2.2). Whereas, the economic variable X_t follows the same dynamics in (1.2.1) and is a measure of the current economic state, so that a rise in the economy will result in an increase in the value of Y_t. Whereas a downturn in the economy, will lead to a decrease in the value of Y_t.

The gain function f at the time, τ, of launch is usually a function of the current value of the economic project Y_t, and some extra uncertainty that depends on the economic state. A typical example of this is if we have

$$f_\tau = Y_\tau \cdot (\beta_0 + \beta_1 X_\tau + a_0\delta),\ \delta \sim N(0,1). \tag{1.2.7}$$

$\beta_1 X_\tau$ in (1.2.7) above models the profit multiplier depending on the state of the economy and the anticipated profit is given as

$$f(\tau, X_\tau, Y_\tau) = Y_\tau(\beta_0 + \beta_1 X_\tau).$$

where $\beta_i \in \mathbb{R}$, $i = 0, 1$.

2. Put-option under Stochastic Convenience Yield

It is normal that to every additional asset accrued by a person there would be an associated gain and risk. Convenience yield models hypothesizes that the drift of the asset price under the pricing measure $\mathbb{P}$ is also a stochastic process i.e.

$$dY_t = Y_t(X_t\ dt + \sigma_Y\ dB_t), \tag{1.2.8}$$
$$dX_t = \lambda X_t\ dt + \gamma X_t\ dB_t + \sigma_X X_t\ dL_t. \tag{1.2.9}$$

where λ denotes the market price of convenience yield. In the light of this, one can then analyse the pricing of a European put option on an asset Y with maturity T and strike price K i.e.

$$\sup_{\tau \in T} \mathbb{E}[\ e^{-r\tau}(K - Y_\tau)_+\ ],$$

Where the convenience yield is unobservable and has to be obtained by some filtering techniques.

2. Main Results

2.1 Kalman-Bucy Filter for Square Integrable Lévy Processes

2.1.1 Introduction. Filters for Lévy processes can be constructed in myriad of ways. In this section, we shall discuss the filtering problem of square integrable Lévy processes. While this is mathematically challenging in its own right, it also has some important computational advantages over other approaches.

For the filtering problem in the context of square integrable processes, the unobservable process is given by the dynamics

$$dX_t = C_t X_t dt + B_t dL_t^2 + D_t dL_t^1. \tag{2.1.1}$$

Where L^1 is a zero mean square integrable Lévy process defined on Ω and taking values in $\mathbb{R}^{s_2}$. Also, $C,\ B$ and D are locally bounded left continuous $s_1 \times s_1,\ s_2 \times b$ and $s_1 \times a$ [1] matrix-valued functions.

In a similar fashion, for $s_2 \leq s_1$, the observed process Y_t taking values in $\mathbb{R}^{s_1}$ is given by the dynamics

$$dY_t = A_t X_t dt + B_t dL_t^2, \quad 0 \leq t \leq T. \tag{2.1.2}$$

Where L_t^2 is a square integrable Lévy process defined on Ω and taking values in $\mathbb{R}^b$. L_t^1 and L_t^2 are independent and A and B are locally bounded left continuous $s_2 \times s_1$ and $s_2 \times b$ matrix-valued functions.

The method employed in this section for the construction of the filtering model is a special case of the Kalman-Bucy filtering where the noise at both the observed and unobservable state are processes with orthogonal increments.

We can now proceed with the construction of the linear filtering problem in the setting of square integrable processes.

2.1.2 Filtering Model. Given the probability space $(\Omega, \mathcal{F}, P)$, suppose at time $t \geq 0$, we have the Lévy driven unobservable process X given by (2.1.1), let us suppose that X_0 is $\mathcal{F}_0$ measurable, this condition ensures that (2.1.1) has a unique solution with right continuity and left limits. The explicit solution to (2.1.1) is as derived below.

Suppose

$$W_t = \exp\left(-\int_0^t C_v dv\right) X_t. \tag{2.1.3}$$

Then by (2.1.1), we have that,

$$\begin{aligned} dW_t &= -C_t \exp\left(-\int_0^t C_v dv\right) X_t\ dt + \exp\left(-\int_0^t C_v dv\right)\left(C_t X_t dt + B_t dL_t^2 + D_t L_t^1\right) \\ &= \exp\left(-\int_0^t C_v dv\right) B_t dL_t^2 + \exp\left(-\int_0^t C_v dv\right) D_t dL_t^1. \end{aligned}$$

[1] s_1, s_2, a and b belong to the set, $\mathbb{N}$, of Natural numbers.

Integrating over the time interval $[0,t]$, we obtain

$$\begin{aligned} W_t &= W_0 + \int_0^t \exp\left(-\int_0^s C_v dv\right) B_s dL_s^2 + \int_0^t \exp\left(-\int_0^s C_v dv\right) D_s dL_s^1 \\ &= W_0 + \int_0^t \exp\left(-\int_0^t C_v dv\right) \left(B_s dL_s^2 + D_s dL_s^1\right). \end{aligned}$$

And so from (2.1.3) we have

$$\begin{aligned} X_t &= \exp\left(\int_0^t C_v dv\right) W_t \\ &= \exp\left(\int_0^t C_v dv\right) \left(W_0 + \int_0^t \exp\left(-\int_0^s C_v dv\right)\right) \left(B_s dL_s^2 + D_s dL_s^1\right) \\ X_t &= exp\left(\int_0^t C_v dv\right) X_0 + \exp\left(\int_0^t C_v dv\right) \left(\int_0^t \exp\left(-\int_0^s C_v dv\right) \left(B_s dL_s^2 + D_s dL_s^1\right)\right) \\ X_t &= \exp\left(\int_0^t C_v dv\right) X_0 + \int_0^t \exp\left(\int_s^t C_v dv\right) \left(B_s dL_s^2 + D_s dL_s^1\right). \end{aligned} \tag{2.1.4}$$

2.1.3 Definition. Suppose $(G(t),\ t \in [0,T])$ is any square integrable process with values in $\mathbb{R}^{s_2}$ such that $\mathbb{E}[\mid G_t \mid^2] < \infty \ \ \forall \ \ t \geq 0$, then we define $\mathcal{L}(G,T)$ to be the closure in the space, $L^2(\Omega, \mathcal{F}, P; \mathbb{R}^{s_1})$, of square integrable functions of all finite linear combinations;

$$J_0 + J_1 G_{t_1} + \cdots + J_k G_{t_k}; \quad 0 \leq t_1 < \cdots < t_n \leq T, J_0 \in \mathbb{R}^{s1},$$

and $J_i,\ i = 1:n$ are arbitrary $s_1 \times s_1$ matrices.

The observation process defined in (2.1.2) with initial condition $Y_0 = 0$ has a unique cádlag solution $Y = Y_t,\ t \geq 0$, taking values in $\mathbb{R}^{s_2}$. Since Y is also square integrable, we have that,

$$\mathbb{E}(|Y_t|^2) < \infty \quad \forall \quad t \geq 0.$$

In what follows, we would limit our work to the time interval $[0,T]$ and make the assumption that $B_t {B_t}^T$ is non-singular for all $t \in [0,T]$. We shall also assume that the map $t \to B_t {B_t}^T$ is bounded away from zero. We define

$$H_t := \left(B_t {B_t}^T\right)^{-\frac{1}{2}} \quad \text{for all } t \in [0,T].$$

2.1.4 Remark. Given a Hilbert space $\mathcal{H}$, suppose K is an orthogonal projection in $\mathcal{H}$, then we write;

$$K^{\perp} = I - K,$$

where I denotes the identity operator.

From now on, we would designate the orthogonal projection from $L^2(\Omega, \mathcal{F}, P; \mathbb{R}^{s_1})$ onto $\mathcal{L}(W,T)$ by $Q_{\mathcal{L}}$. Thus, for all $t \in [0,T]$, we can define the orthogonal projection of X_t onto $\mathcal{L}(W,T)$ as

$$\hat{X}_t = Q_{\mathcal{L}}(X_t), \tag{2.1.5}$$

where $\hat{X}_t$ is taken to be the best linear estimate of X_t. From (2.1.5), we deduce that,

$$X_t - \hat{X}_t = (I - Q_{\mathcal{L}})X_t. \tag{2.1.6}$$

So that from (2.1.6) and Definition 2.1.3 above, we see that,

$$X_t - \hat{X}_t \text{ is orthogonal to } \mathcal{L}(W,T).$$

2.1.5 Lemma. For all $t \in [0, T]$, we have that;

$$\mathbb{E}(\hat{X}_t) = \mathbb{E}(X_t),$$

with $\hat{X}_0 := \mathbb{E}(X_0)$.

Proof. Recall that expected values define a real inner product i.e.

$$\mathbb{E}[XY] = \langle XY \rangle,$$

for random variables X and Y.

Thus,

$$\begin{aligned}
\mathbb{E}[\hat{X}] &= \mathbb{E}[\hat{X}_t \cdot 1] \\
&= \left\langle \hat{X}_t, 1 \right\rangle \\
&= \langle Q_{\mathcal{L}}(X_t), 1 \rangle \quad \text{(by (2.1.5))} \\
&= \langle X_t, Q_{\mathcal{L}}(1) \rangle \quad \text{(by a property of orthogonal projection)} \\
&= \langle X_t, 1 \rangle \\
&= \mathbb{E}[X_t \cdot 1] = \mathbb{E}[X_t] \quad \forall \quad t \in [0, T].
\end{aligned}$$

□

2.2 Innovation Process

The innovation process $R = (R_t,\ t \in [0, t]),\ R \in \mathbb{R}^{s_2}$ with orthogonal increments, can be defined as

$$R_t = Y_t - \int_0^t A_s \hat{X}_s ds. \tag{2.2.1}$$

Using (2.1.2), we have that,

$$dR_t = A_t(X_t - \hat{X}_t)dt + B_t dL_t^2. \tag{2.2.2}$$

We shall now state without proof a theorem and a corollary due to (Blackwood, 2014) which we would refer to at a later stage.

2.2.1 Theorem. *A Lévy process is a martingale if and only if it takes the form*

$$X(t) = B_A(t) + \int_{\mathbb{R}^d - \{0\}} x\tilde{N}(t, dx).$$

2.2.2 Corollary. A Lévy process possesses mean zero if and only if it is a martingale.

2.2.3 Proposition. Suppose we define the process $Z = \{Z_t\}_{t \in [0,T]}$ taking values in $\mathbb{R}^{s_2}$ to be the unique solution of

$$dZ_t = H_t\, dR_t, \tag{2.2.3}$$

with $Z_0 = 0$ almost surely. Then we claim that,

1. Z_t has orthogonal increments.

2. Z_t has zero expectation and,

3. $\forall\ 0 \leq s < t,\ t < T,\ \mathbb{E}[Z_s {Z_t}^T] = \Psi_s^2$. where;

$$\Psi_t^2 := \int_0^t H_s B_s \gamma_2 {B_s}^T {H_s}^T \ ds \ \text{ and } \ \gamma_2 = \sigma_2 {\sigma_2}^T$$

We shall assume the existence of $\Psi_t^{2^{-1}}$ for all $t \in [0, T]$ and the boundedness of the map $t \to \Psi_t^{2^{-1}}$ (see (Applebaum and Blackwood, 2015)).

Proof. 1. Since R_t has orthogonal increments, then it follows that Z_t also has orthogonal increments. □

Proof. 2.

$$\begin{aligned}\mathbb{E}[Z_t] &= \mathbb{E}\left[\int_0^t H_s \ dR_s\right] \\ &= \mathbb{E}\left[\int_0^t H_s \left[A_s(X_s - \hat{X}_s)ds + B_s dL_s^2\right]\right],\end{aligned}$$

Then by Fubinni's Theorem, we obtain

$$\begin{aligned}\mathbb{E}[Z_t] &= \int_0^t \left[H_s A_s \ \mathbb{E}(X_s - \hat{X}_s)\right] ds + \mathbb{E}\left[\int_0^t H_s B_s \ dL_s^2\right] \\ &= 0.\end{aligned}$$

The above holds from the fact that the expectation of a stochastic integral is zero with respect to a square integrable martingale and also from the fact that,

$$\begin{aligned}\mathbb{E}[X_s - \hat{X}_s] &= \mathbb{E}[X_s] - \mathbb{E}[\hat{X}_s] \\ &= \mathbb{E}[X_s] - \mathbb{E}[\mathbb{E}(\hat{X}_s)] \\ &= 0.\end{aligned}$$

□

Proof. 3. Using orthogonal increments and the fact that $\mathbb{E}[Z_t] = 0$, we have

$$\begin{aligned}\mathbb{E}[Z_s {Z_t}^T] &= \mathbb{E}[Z_s {Z_t}^T - Z_t {Z_t}^T + Z_t {Z_t}^T] \\ &= \mathbb{E}[(Z_s - Z_t){Z_t}^T + Z_t {Z_t}^T] \\ &= \mathbb{E}\left[Z_s - Z_t\right] \ \mathbb{E}\left[{Z_t}^T\right] + \mathbb{E}\left[Z_t {Z_t}^T\right] \\ &= \mathbb{E}[Z_t {Z_t}^T].\end{aligned}$$

Employing Ito's product formula, we have that,

$$d[Z_t \ {Z_t}^T] = Z_t \ d{Z_t}^T + dZ_t \ {Z_t}^T + dZ_t \ d{Z_t}^T. \tag{2.2.4}$$

Let us now evaluate $dZ_t \ d{Z_t}^T$ separately. This is done as follows,

$$\begin{aligned}dZ_t \ d{Z_t}^T &= [H_t \ dR_t] \ [H_t \ dR_t]^T = H_t \ dR_t \ d{R_t}^T {H_t}^T \\ &= H_t[A_t(X_t - \hat{X}_t)dt + B_t \ dL_t^2] \ [A_t(X_t - \hat{X}_t)dt + B_t \ dL_t^2]^T {H_t}^T \\ &= H_t[A_t(X_t - \hat{X}_t)dt + B_t \ dL_t^2] \ [(A_t(X_t - \hat{X}_t))^T dt + dL_t^{2^T} \ {B_t}^T]{H_t}^T.\end{aligned}$$

But $(dL_t^2)^2 \sim dt$ and dt is very small such that $(dt)^2$ is negligible. This implies that,

$$dZ_t\ dZ_t{}^T = H_t\ B_t\ dL_t^2\ dL_t^{2^T}\ B_t{}^T H_t{}^T. \tag{2.2.5}$$

□

Recall that for a square integrable zero mean Lévy process, we have that,

$$X_t^i = B_t^i + \int_{\mathbb{R}_0} x\tilde{N}_i(t, dx), \tag{2.2.6}$$

where $B_t^i = \sigma_i B_t$ and B_t is a standard Brownian motion. Equation (2.2.6) then implies that,

$$dX_t^i = \sigma_i\ dB_t^i + \int_{\mathbb{R}_0} x\tilde{N}_i(dt, dx). \tag{2.2.7}$$

Using Equation (2.2.6), we have that,

$$dL_t^2 = \sigma_2\ dB_t^2 + \int_{\mathbb{R}_0} xN(dt, dx). \tag{2.2.8}$$

Therefore,

$$\begin{aligned} dL_t^2\ dL_t^{2^T} &= \left[\sigma_2\ dB_t^2 + \int_{\mathbb{R}_0} xN(dt, dx)\right]\left[dB_t^{2^T}\ \sigma_2{}^T + \int_{\mathbb{R}_0} x^T N(dt, dx)\right] \\ &= \sigma_2\ dB_t^2\ dB_t^{2^T}\ \sigma_2{}^T + \left(\int_{\mathbb{R}_0} xx^T N(dt, dx)\right). \end{aligned} \tag{2.2.9}$$

Thus;

$$dZ_t\ dZ_t{}^T = H_t B_t \left[\sigma_2 \sigma_2{}^T dt + \left(\int_{\mathbb{R}_0} xx^T N(dt, dx)\right)\right] B_t{}^T H_t{}^T.$$

This implies from (2.2.4) that we have

$$d\left(Z_t\ Z_t{}^T\right) = Z_t\ dZ_t{}^T + dZ_t\ Z_t{}^T + H_t B_t \left[\sigma_2 \sigma_2{}^T dt + \left(\int_{\mathbb{R}_0} xx^T\ N(dt, dx)\right)\right] B_t{}^T H_t{}^T,$$

by using Theorem 2.2.1 and Corollary 2.2.2.

Therefore,

$$\begin{aligned} d\left(Z_t\ Z_t^T\right) &= Z_t \left(H_t\ dR_t\right)^T + \left(H_t\ dR_t\right) Z_t{}^T + H_t B_t \left[\sigma_2 \sigma_2{}^T dt + \left(\int_{\mathbb{R}_0} xx^T\ N(dt, dx)\right)\right] B_t{}^T H_t{}^T \\ &= Z_t\ dR_t^T\ H_t^T + H_t\ dR_t\ Z_t{}^T + H_t B_t \left[\sigma_2 \sigma_2{}^T dt + \left(\int_{\mathbb{R}_0} xx^T\ N(dt, dx)\right)\right] B_t{}^T H_t{}^T. \end{aligned}$$

By using (2.2.2) we have that,

$$\begin{aligned} d\left(Z_t Z_t{}^T\right) = Z_t &\left[A_t(X_t - \hat{X}_t)\ dt + B_t\ dL_t^2\right]^T H_t{}^T + H_t \left[A_t(X_t - \hat{X}_t)\ dt + B_t\ dL_t^2\right] Z_t{}^T \\ &+ H_t B_t \left[\sigma_2 \sigma_2{}^T dt + \left(\int_{\mathbb{R}_0} xx^T\ N(dt, dx)\right)\right] B_t{}^T H_t{}^T. \end{aligned}$$

Using the fact that we have no jump, we now take the integral and expectation of both sides and apply Fubinni's Theorem, to obtain

$$\begin{aligned}\mathbb{E}[Z_t\ Z_t^T] &= Z_0{Z_0}^T + \int_0^t Z_s\ \mathbb{E}(X_s - \hat{X}_s)^T\ {A_s}^T H_s^T\ ds + \mathbb{E}\left(\int_0^t Z_s\ d{L_s^2}^T {B_s}^T {H_s}^T\right) \\ &\quad + \int_0^t H_s\ A_s\ \mathbb{E}(X_s - \hat{X}_s)\ {Z_s}^T\ ds + \mathbb{E}\left(\int_0^t H_s\ B_s\ dL_s^2 {Z_s}^T\right) + \mathbb{E}\left(\int_0^t H_s\ B_s\ \gamma_2\ {B_s}^T {H_s}^T\ ds\right) \\ &= Z_0\ {Z_0}^T + \int_0^t H_s\ B_s\ \gamma_2\ {B_s}^T {H_s}^T ds \\ &= \int_0^t H_s\ B_s\ \gamma_2\ {B_s}^T {H_s}^T ds \quad \text{(since } Z_0 = 0 \text{ almost surely)} \\ &= \Psi_t^2,\end{aligned}$$

where $\gamma_2 = \sigma_2{\sigma_2}^T$

The above was obtained by using the latter part of the second proof and the fact that the expectation of a stochastic integral with respect to a martingale is zero.

Using the ideas in (Applebaum and Blackwood, 2015), we shall suppose that Ψ_t^2 is invertible, and that the map $t \to (\Psi_t^2)^{-1}$ is bounded. We now proceed with the proof of an important theorem that gives the differential equation satisfied by the variance E_t.

2.2.4 Theorem. *Let $t \in [0, T]$ and let the mean-squared error be defined as*

$$E_t = \mathbb{E}\left[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T\right]. \tag{2.2.10}$$

Then the differential equation satisfied by E_t is given as

$$\frac{dE_t}{dt} = (B_t\gamma_2{B_t}^T + D_t\gamma_1{D_t}^T) - \left(E_t\ {A_t}^T H_t^T \left(\Psi_t^2\right)^{-1} H_t\ A_t\ {E_t}^T\right) + C_tE_t + E_t{C_t}^T,$$

with associated initial condition $E_0 = cov(X_0)$.

Proof. To do this, we shall employ lemma 2.3.8 of (Blackwood, 2014) from which we have that,

$$\hat{X}_t = \mathbb{E}[X_t] + \int_0^t g(s,t)\ dZ_s, \tag{2.2.11}$$

where,

$$g(s,t) = \frac{\partial}{\partial s}\mathbb{E}[X_t\ {Z_s}^T]\ (\Psi_s^2)^{-1}.$$

From Lemma 2.4.1 of (Blackwood, 2014), we have that,

$$g(s,t) = \exp\left(\int_s^t C_v dv\right) E_s\ {A_s}^T {H_s}^T (\Psi_s^2)^{-1}\ \ \forall\ \ 0 \le t \le T. \tag{2.2.12}$$

Also from (2.2.10) we have,

$$\begin{aligned}E_t &= \mathbb{E}\left[(X_t - \hat{X}_t)(X_t - \hat{X}_t)^T\right] \\ &= \mathbb{E}\left[(X_t - \hat{X}_t)({X_t}^T - \hat{X}_t^T)\right] \\ &= \mathbb{E}\left[X_t{X_t}^T\right] - \mathbb{E}\left[\hat{X}_t\hat{X}_t^T\right].\end{aligned}$$

Using (2.2.11), we obtain

$$\begin{aligned} E_t &= \mathbb{E}[X_t {X_t}^T] - \mathbb{E}\left[\left(\mathbb{E}[X_t] + \int_0^t g(s,t)\ dZ_s\right)\left(\mathbb{E}[X_t] + \int_0^t\ g(s,t)\ dZ_s\right)^T\right] \\ &= \mathbb{E}[X_t {X_t}^T] - \mathbb{E}\left[\left(\mathbb{E}[X_t] + \int_0^t g(s,t)\ dZ_s\right)\left(\mathbb{E}[X_t]^T + \int_0^t\ d{Z_s}^T g(s,t)^T\right)\right]. \end{aligned}$$

Using Ito's isometry and the zero expectation of the stochastic integral, we obtain

$$\begin{aligned} E_t &= \mathbb{E}\left[X_t {X_t}^T\right] - \mathbb{E}[X_t]\mathbb{E}[X_t]^T - \mathbb{E}\left(\int_0^t g(s,t)\ dZ_s\ \left(\int_0^t\ g(s,t)\ dZ_s\right)^T\right) \\ &= \mathbb{E}\left[X_t {X_t}^T\right] - \mathbb{E}[X_t]\mathbb{E}[X_t]^T - \mathbb{E}\left(\int_0^t g(s,t)\ dZ_s\ \int_0^t\ d{Z_s}^T\ g(s,t)^T\right). \end{aligned}$$

Using (2.2.5), we have

$$\begin{aligned} E_t &= \mathbb{E}\left[X_t {X_t}^T\right] - \mathbb{E}[X_t]\mathbb{E}[X_t]^T - \mathbb{E}\left(\int_0^t g(s,t)\ \left(\int_0^t H_s\ B_s\ dL_s^2\ {dL_s^2}^T {B_s}^T {H_s}^T\right)\ g(s,t)^T\ ds\right) \\ &= \mathbb{E}\left[X_t {X_t}^T\right] - \mathbb{E}[X_t]\mathbb{E}[X_t]^T - \mathbb{E}\left(\int_0^t g(s,t)\ \left(\int_0^t H_s\ B_s\ \gamma_2\ {B_s}^T\ {H_s}^T\ ds\right)\ g(s,t)^T\ ds\right) \quad \text{[by using (2.2.9)]} \\ &= \mathbb{E}\left[X_t {X_t}^T\right] - \mathbb{E}[X_t]\mathbb{E}[X_t]^T - \int_0^t g(s,t)\ \left(\int_0^t H_s\ B_s\ \gamma_2\ {B_s}^T {H_s}^T\ ds\right)\ g(s,t)^T\ ds \\ &= \mathbb{E}\left[X_t {X_t}^T\right] - \mathbb{E}[X_t]\mathbb{E}[X_t]^T - \int_0^t g(s,t)\ \Psi_s^2\ g(s,t)^T\ ds. \end{aligned}$$

This implies that,

$$E_t = J_t - \mathbb{E}[X_t]\mathbb{E}[X_t]^T - \int_0^t g(s,t)\ \Psi_s^2\ g(s,t)^T\ ds, \tag{2.2.13}$$

where, $J_t = \mathbb{E}[X_t {X_t}^T]$.

Differentiating both sides of (2.2.13), we obtain

$$\begin{aligned} \frac{dE_t}{dt} = \frac{dJ_t}{dt} &- \frac{d}{dt}\mathbb{E}[X_t]\mathbb{E}[X_t]^T - g(t,t)\ \Psi_t^2\ g(t,t)^T \\ &- \int_0^t g(s,t)\ \Psi_t^2\ \frac{\partial}{\partial t}\ g(s,t)^T\ ds - \int_0^t \frac{\partial}{\partial t}\ g(s,t)\ \Psi_t^2\ g(s,t)^T\ ds. \end{aligned} \tag{2.2.14}$$

Using the expression of X_t in (2.1.4), we obtain

$$
\begin{aligned}
J_t &= \mathbb{E}\left[X_t X_t^T\right] \\
&= \mathbb{E}\Big[\left[\exp\left(\int_0^t C_v\, dv\right) X_0 + \int_0^t \exp\left(\int_s^t C_v\, d_v\right)\left(B_s\, dL_s^2 + D_s\, dL_t^1\right)\right] \\
&\qquad\qquad \times \\
&\quad \left[\exp\left(\int_0^t C_v\, dv\right) X_0 + \int_0^t \exp\left(\int_s^t C_v\, d_v\right)\left(B_s\, dL_s^2 + D_s\, dL_t^1\right)\right]^T\Big] \\
&= \mathbb{E}\Big[\left[\exp\left(\int_0^t C_v\, dv\right) X_0 + \int_0^t \exp\left(\int_s^t C_v\, d_v\right)\left(B_s\, dL_s^2 + D_s\, dL_t^1\right)\right] \\
&\qquad\qquad \times \\
&\quad \left[X_0{}^T \exp\left(\int_0^t C_v{}^T\, dv\right) + \int_0^t \left(dL_s^{2T}\, B_s{}^T + dL_t^{1T}\, D_s{}^T\right)\exp\left(\int_s^t C_v{}^T\, d_v\right)\right]\Big] \\
&= \exp\left(\int_0^t C_v\, dv\right)\mathbb{E}[X_0 X_0^T]\ \exp\left(\int_0^t C_v{}^T\, dv\right) + \int_0^t \exp\left(\int_s^t C_v\, d_v\right)\left(B_s\, dL_s^2 + D_s\, dL_s^1\right) \\
&\qquad \left(dL_s^{2T}\, B_s{}^T + dL_s^{1T}\, D_s{}^T\right)\exp\left(\int_s^t C_v{}^T\, dv\right) \\
&= \exp\left(\int_0^t C_v\, dv\right)\mathbb{E}[X_0 X_0{}^T]\ \exp\left(\int_s^t C_v{}^T\, dv\right) \\
&\qquad + \int_0^t \exp\left(\int_s^t C_v\, dv\right)(B_s\, \gamma_2\, B_s{}^T + D_s\, \gamma_1\, D_s{}^T)\exp\left(\int_s^t C_v{}^T\, dv\right) ds. \qquad (2.2.15)
\end{aligned}
$$

Differentiating both sides of (2.2.15) gives

$$
\begin{aligned}
\frac{dJ_t}{dt} &= C_t \exp\left(\int_0^t C_v\, dv\right)\mathbb{E}[X_0 X_0{}^T]\ \exp\left(\int_0^t C_v{}^T\, dv\right) \qquad (2.2.16) \\
&\quad + \exp\left(\int_0^t C_v\, dv\right)\mathbb{E}[X_0 X_0{}^T]\ \exp\left(\int_0^t C_v{}^T\, dv\right) C_t{}^T \\
&\quad + (B_t\, \gamma_2\, B_t{}^T + D_t\, \gamma_1\, D_t{}^T) \\
&\quad + \int_0^t C_t \exp\left(\int_s^t C_v\, dv\right)(B_s\, \gamma_2\, B_s{}^T + D_s\, \gamma_1\, D_s{}^T)\exp\left(\int_s^t C_v{}^T\, dv\right)\ ds \\
&\quad + \int_0^t \exp\left(\int_s^t C_v\, dv\right)(B_s\, \gamma_2\, B_s{}^T + D_s\, \gamma_1\, D_s{}^T)\exp\left(\int_s^t C_v{}^T\, dv\right) C_t^T\, ds. \\
\Longrightarrow \frac{dJ_t}{dt} &= C_t\Big[\exp\left(\int_0^t C_v\, dv\right)\mathbb{E}\left[X_0 X_0{}^T\right]\exp\left(\int_0^t C_v{}^T\, dv\right) \\
&\quad + \int_0^t \exp\left(\int_s^t C_v\, dv\right)\ (B_s\, \gamma_2\, B_s{}^T + D_s\, \gamma_1\, D_s{}^T)\exp\left(\int_s^t C_v{}^T\, dv\right) ds\Big] \\
&\quad + (B_t\, \gamma_2\, B_t{}^T + D_t\, \gamma_1\, D_t{}^T) \\
&\quad + \Big[\exp\left(\int_0^t C_v\, dv\right)\mathbb{E}\left[X_0 X_0{}^T\right]\exp\left(\int_0^t C_v{}^T\, dv\right) \\
&\quad + \int_0^t \exp\left(\int_s^t C_v\, dv\right)\ (B_s\, \gamma_2\, B_s{}^T + D_s\, \gamma_1\, D_s{}^T)\exp\left(\int_s^t C_v{}^T\, dv\right) ds\Big] C_t^T \\
&= (B_t\, \gamma_2\, B_t{}^T + D_t\, \gamma_1\, D_t{}^T) + C_t\, J_t + J_t\, C_t{}^T. \qquad (2.2.17)
\end{aligned}
$$

In a similar fashion, let us differentiate $\mathbb{E}(X_t)\mathbb{E}(X_t)^T$ as follows,

$$\begin{aligned}
\mathbb{E}(X_t)\mathbb{E}(X_t)^T &= \mathbb{E}\left[\exp\left(\int_0^t C_v\ dv\right) X_0 + \int_0^t \exp\left(\int_s^t C_v\ d_v\right)\left(B_s\ dL_s^2 + D_s\ dL_s^1\right)\right] \\
&\times \\
&\mathbb{E}\left[\exp\left(\int_0^t C_v\ dv\right) X_0 + \int_0^t \exp\left(\int_s^t C_v\ d_v\right)\left(B_s\ dL_s^2 + D_s\ dL_s^1\right)\right]^T \\
&= \mathbb{E}\left[\exp\left(\int_0^t C_v\ dv\right) X_0 + \int_0^t \exp\left(\int_s^t C_v\ d_v\right)\left(B_s\ dL_s^2 + D_s\ dL_s^1\right)\right] \\
&\times \\
&\mathbb{E}\left[{X_0}^T \exp\left(\int_0^t {C_v}^T\ dv\right) + \int_0^t \left(dL_s^{2T}\ {B_s}^T + dL_s^{1T}\ {D_s}^T\right)\exp\left(\int_s^t {C_v}^T\ d_v\right)\right]
\end{aligned}$$

But the expectation of a stochastic integral with respect to a martingale is zero. This implies that,

$$\begin{aligned}
\mathbb{E}(X_t)\mathbb{E}(X_t)^T &= \mathbb{E}\left[\exp\left(\int_0^t C_v\ dv\right) X_0\right]\mathbb{E}\left[{X_0}^T \exp\left(\int_0^t {C_v}^T\ dv\right)\right] \\
&= \mathbb{E}\left[\exp\left(\int_0^t C_v\ dv\right) X_0{X_0}^T \exp\left(\int_0^t {C_v}^T\ dv\right)\right] \\
\implies \frac{d}{dt}\mathbb{E}(X_t)\mathbb{E}(X_t)^T &= C_t\ \mathbb{E}\left[\exp\left(\int_0^t C_v\ dv\right) X_0{X_0}^T \exp\left(\int_0^t {C_v}^T\ dv\right)\right] \\
&+ \mathbb{E}\left[\exp\left(\int_0^t C_v\ dv\right) X_0{X_0}^T \exp\left(\int_0^t {C_v}^T\ dv\right)\right] {C_t}^T \\
&= C_t\ \mathbb{E}(X_t)\ \mathbb{E}(X_t)^T + \mathbb{E}(X_t)\ \mathbb{E}(X_t)^T {C_t}^T. \qquad (2.2.18)
\end{aligned}$$

Now substituting (2.2.13) and (2.2.17) into (2.2.18), we obtain

$$\begin{aligned}
\frac{dE_t}{dt} &= (B_t\ \gamma_2\ {B_t}^T + D_t\ \gamma_1\ {D_t}^T) + C_tJ_t + J_t{C_t}^T - \left[E_t\ {A_t}^T\ {H_t}^T(\Psi_t^2)^{-1}\right]\Psi_t^2\left[\left(\Psi_t^2\right)^{-1} H_t\ A_t\ {E_t}^T\right] \\
&- C_t\int_0^t g(s,t)\Psi_s^2 g(s,t)^T\ ds - \int_0^t g(s,t)\ \Psi_s^2\ g(s,t)^T\ ds\ {C_t}^T - C_t\ \mathbb{E}[X_t]\mathbb{E}[X_t]^T - \mathbb{E}[X_t]\mathbb{E}[X_t]^T\ {C_t}^T \\
&= (B_t\ \gamma_2\ {B_t}^T + D_t\ \gamma_1\ {D_t}^T) - \left(E_t\ {A_t}^T{H_t}^T\left(\Psi_t^2\right)^{-1} H_tA_t{E_t}^T\right) + C_tJ_t - C_t\int_0^t g(s,t)\ \Psi_s^2\ g(s,t)^T\ ds \\
&- C_t\mathbb{E}[X_t]\mathbb{E}[X_t]^T + J_t{C_t}^T - \int_0^t g(s,t)\ \Psi_s^2\ g(s,t)^T\ ds\ {C_t}^T - \mathbb{E}(X_t)\mathbb{E}[X_t]^T\ {C_t}^T.
\end{aligned}$$

Using (2.2.13), this becomes;

$$\begin{aligned}
\frac{dE_t}{dt} &= (B_t\ \gamma_2\ {B_t}^T + D_t\ \gamma_1\ {D_t}^T) - \left(E_t\ {A_t}^T{H_t}^T\left(\Psi_t^2\right)^{-1} H_t\ A_t\ {E_t}^T\right) \\
&+ C_t\left(E_t + \mathbb{E}[X_t]\mathbb{E}[X_t]^T + \int_0^t g(s,t)\Psi_s^2\ g(s,t)^T\ ds\right) \\
&- C_t\int_0^t g(s,t)\ \Psi_s^2\ g(s,t)^T\ ds - C_t\mathbb{E}[X_t]\mathbb{E}[X_t]^T \\
&+ \left(E_t + \mathbb{E}[X_t]\mathbb{E}[X_t]^T + \int_0^t g(s,t)\ \Psi_s^2\ g(s,t)^T\ ds\right){C_t}^T \\
&- \int_0^t g(s,t)\ \Psi_s^2\ g(s,t)^T\ ds\ {C_t}^T - \mathbb{E}[X_t]\mathbb{E}(X_t)^T\ {C_t}^T \\
&= (B_t\ \gamma_2\ {B_t}^T + D_t\ \gamma_1\ {D_t}^T) - \left(E_t\ {A_t}^T{H_t}^T\left(\Psi_t^2\right)^{-1} H_t\ A_t\ {E_t}^T\right) + C_tE_t + E_t{C_t}^T,
\end{aligned}$$

with associated initial condition $E_0 = \text{cov}(X_0)$. □

With the stochastic differential equation satisfied by the mean square error E_t, we now proceed to obtain an expression for the derivative of the linear estimator, $\hat{X}$.

2.2.5 Theorem. *The Stochastic differential equation satisfied by the solution*

$$\hat{X}_t = Q_{\mathcal{L}}(X_t),$$

is given as

$$d\hat{X}_t = C_t\hat{X}_t\ dt + E_t\ {A_t}^T {H_t}^T (\Psi_t^2)^{-1}\ H_t \left[dY_t - A_t\hat{X}_t\ dt\right]. \tag{2.2.19}$$

Proof. From (2.2.11), we have that;

$$\hat{X}_t = \mathbb{E}[X_t] + \int_0^t g(s,t)\ dZ_s. \tag{2.2.20}$$

Differentiating both sides of (2.2.20), we obtain;

$$d\hat{X}_t = \frac{d}{dt}\ \mathbb{E}[X_t]\ dt + g(t,t)\ dZ_t + \left(\int_0^t \frac{\partial}{\partial s} g(s,t)\ dZ_s\right)\ dt. \tag{2.2.21}$$

Using (2.1.4) and the fact that the expectation of a stochastic integral with respect to a martingale is zero, we obtain an expression for $\frac{d}{dt}\mathbb{E}[X_t]dt$ as follows,

$$\begin{aligned}\frac{d}{dt}\ \mathbb{E}[X_t]dt &= \frac{d}{dt}\ \mathbb{E}\left[\exp\left(\int_0^t C_v\ dv\right) X_0 + \int_0^t \exp\left(\int_s^t C_v\ dv\right)\left(B_s\ dL_{sT}^2 + D_s\ dL_s^1\ dt\right)\right] dt\\ &= \frac{d}{dt}\ \mathbb{E}\left[\exp\left(\int_0^t C_v\ dv\right) X_0\right] dt\\ &= C_t\ \mathbb{E}[X_t]\ dt.\end{aligned}$$

Using (2.2.12), (2.2.21) becomes

$$\begin{aligned} d\hat{X}_t &= C_t\ \mathbb{E}[X_t]\ dt + E_t\ {A_t}^T\ {H_s}^T \left(\Psi_t^2\right)^{-1}\ dZ_t + C_t\ \left(\int_0^t g(s,t)\ dZ_s\right) dt\\ &= C_t\ \mathbb{E}[X_t]\ dt + E_t\ {A_t}^T\ {H_s}^T \left(\Psi_t^2\right)^{-1}\ dZ_t + C_t\ \left(\hat{X}_t - \mathbb{E}[X_t]\right) dt \quad \text{(by using (2.2.20))}\\ &= C_t\hat{X}_t\ dt + E_t\ {A_t}^T {H_t}^T \left(\Psi_t^2\right)^{-1}\ dZ_t. \end{aligned} \tag{2.2.22}$$

But recall from (2.2.3) that,

$$dZ_t = H_t\ dR_t. \tag{2.2.23}$$

Therefore using (2.2.2), (2.2.23) becomes

$$\begin{aligned} dZ_t &= H_t\left[A_t\left(X_t - \hat{X}_t\right) + B_t\ dL_t^2\right]\\ &= H_t\left[A_tX_t\ dt + B_t\ dL_t^2 - A_t\hat{X}_t\ dt\right]\\ &= H_t\left[dY_t - A_t\hat{X}_t\ dt\right] \quad \text{(using (2.2.1))} \end{aligned} \tag{2.2.24}$$

Substituting (2.2.24) into (2.2.22), we obtain

$$d\hat{X}_t = C_t\hat{X}_t\ dt + E_t\ {A_t}^T {H_t}^T (\Psi_t^2)^{-1}\ H_t \left[dY_t - A_t\hat{X}_t\ dt\right]. \tag{2.2.25}$$

$\hat{X}_t$ is the best linear estimate of X_t given by the Kalman-Bucy filter. □

With the filter Π_t described in Theorems 2.2.4 and 2.2.5 respectively, we can now write the optimal stopping problem under partial information as the optimal stopping problem under full information defined as

$$S_0(\pi_0, y_0) = \sup_{\tau \in \mathcal{T}^Y} \mathbb{E}_{\overline{\mathbb{P}}}[\, F(\Pi_t, Y_t) \mid X_0 \sim \pi_0, Y_0 = y_0]. \tag{2.2.26}$$

In the next section, we shall determine some optimal stopping criterion for (2.2.26) using Snell envelope theory and the dynamic programming principle.

2.3 Snell Envelope and Dynamic Programming Principle

Before we begin with Snell envelope, we shall present a very important result on the essential supremum.

The Essential Supremum

2.3.1 Definition. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $\mathcal{M}$ be a non-empty family of non-negative random variables defined on $(\Omega, \mathcal{F}, \mathbb{P})$. Then, the essential supremum of $\mathcal{M}$, denoted by $ess\ \sup\ \mathcal{M}$, is a random variable X^* satisfying:

1. $X \leq X^*$ almost surely, for all $X \in \mathcal{M}$, and
2. if Y is a random variable satisfying $X \leq Y$ almost surely for all $X \in \mathcal{M}$, then $X^* \leq Y$ almost surely (see (Karatzas and Shreve, 1998)).

We shall state without proof a proposition and two resulting remarks that are fundamental to our work.

2.3.2 Proposition. The essential supremum of a set G, always exists and there exists a sequence, $h_n \in G$ such that,

$$ess\ \sup(G) = \sup_n h_n \text{ almost surely.}$$

2.3.3 Remark. 1. Essential supremum is always unique up to almost sure equivalence.

2.3.4 Remark. 2. If the set G is directed upwards[2] so that for every $h_1, h_2 \in G$, there exists $h \in G$ such that $h \geq h_1 \vee h_2$ almost surely, then one can claim that there exists a non-decreasing sequence $\{h_n\} \subset G$ such that,

$$ess\ \sup(G) = \lim_n \uparrow\ h_n.$$

We now introduce Snell envelope.

2.3.5 Snell Envelope. Snell envelope is a very vital tool for solving optimal stopping problems and it has great applications in finance. A formal definition of the Snell envelope is given below.

2.3.6 Definition. Given $(\Omega, \mathcal{F}, \mathbb{P})$ to be a complete probability space and $Y = \{Y_n;\ n = 0, 1, \dots\}$ to be a sequence of integrable functions adapted to the filtration $\mathcal{F}$ with the condition that $\sup_n\ {Y_n}^+$ is integrable, then the snell envelope is defined as the process $Z := \{Z_n\}_{n \geq 0}$ given by

$$Z_n = ess \sup_{n \leq \tau \leq T} \mathbb{E}[Y_\tau | \mathcal{F}_n].$$

[2] An upward directed set is a non.empty set G together with reflexive and transitive binary relation $\leq$ and with the additional property that every pair of elements in G is bounded above i.e if $g1, g2 \in G$ then there exists $g \in G$ such that $g1 \leq g$ and $g2 \leq g$.

In the context of our work, the snell envelope of the optimal stopping problem defined in Equation (2.2.26) is given as

$$Z_n = ess \sup_{n \leq \tau \leq T} \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)|\mathcal{F}_n{}^Y].$$

If we define

$$\hat{\mathcal{F}}_n := \mathcal{F}_n{}^Y,$$

then Z_n can be re-written in the form

$$Z_n = ess \sup_{n \leq \tau \leq T} \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau,\ Y_\tau)|\hat{\mathcal{F}}_n].$$

Consider the theorem below.

2.3.7 Theorem. *The process Z_n satisfies the dynamic programming principle;*

$$Z_n = \max\left(F(\Pi_n, Y_n),\ \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{n+1}, Y_{n+1})|\hat{\mathcal{F}}_n]\right), \tag{2.3.1}$$

and,

$$V_n := \sup_{n \leq \tau \leq T} \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)] = \mathbb{E}(Z_n),$$

where V_n is the value function. For every $n \geq 0$, Z is the smallest supermartingale that dominates $F(\Pi_n, Y_n)$.

N.B The expression $\mathbb{E}_{\overline{\mathbb{P}}}[Z_{n+1}|\hat{\mathcal{F}}_n] = \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{n+1}, Y_{n+1})|\hat{\mathcal{F}}_n]$ is called the continuation value.

Proof. To show that $Z_n = \max\{F[\Pi_n, Y_n],\ \mathbb{E}_{\overline{\mathbb{P}}}[Z_{n+1}|\hat{\mathcal{F}}_n]\}$, we need to show that;

$$Z_n \leq \max\{F[\Pi_n, Y_n],\ \mathbb{E}_{\overline{\mathbb{P}}}[Z_{n+1}|\hat{\mathcal{F}}_n]\},$$

and,

$$Z_n \geq \max\{F[\Pi_n, Y_n],\ \mathbb{E}|_{\overline{\mathbb{P}}}\hat{\mathcal{F}}_n[Z_{n+1}]\}.$$

This is done by using the forward and backward proofs as shown below.

Forward Proof.

The process Z_n is $\hat{\mathcal{F}}_n$ measurable and bounded by definition. Furthermore, it satisfies

$$F(\Pi_n, Y_n) \leq Z_n. \tag{2.3.2}$$

We assert that the collection of random variables $\{\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)|\hat{\mathcal{F}}_n] : \tau \geq n\}$ are directed upwards. Suppose τ_1 and τ_2 are two stopping times[3] and suppose we define

$$I := \{\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_1}, Y_{\tau_1})|\hat{\mathcal{F}}_n] \geq \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_2}, Y_{\tau_2})|\hat{\mathcal{F}}_n]\}.$$

Then we have that $I \in \hat{\mathcal{F}}_n$ and the random time τ defined as

$$\tau := \tau_1 \cdot \mathbf{1}_I + \tau_2 \cdot \mathbf{1}_{I^{\complement}}, \tag{2.3.3}$$

[3]A stopping time $\tau : \Omega \to \mathbb{R}_+ \cup \{\infty\}$ is a random time such that $\tau \leq t \in \mathcal{F}_t$ for each $t \geq 0$. Stopping times are usually adapted to the underlying filtration.

is a stopping time. It follows that,

$$\begin{aligned}\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)|\hat{\mathcal{F}}_n] &= \mathbf{1}_I \ \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_1}, Y_{\tau_1})|\hat{\mathcal{F}}_n] + \mathbf{1}_{I^c} \ \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_2}, Y_{\tau_2})|\hat{\mathcal{F}}_n] \\ &= \max\{\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_1}, Y_{\tau_1})|\hat{\mathcal{F}}_n)], \ \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_2}, Y_{\tau_2})|\hat{\mathcal{F}}_n)]\}. \end{aligned} \tag{2.3.4}$$

Then by Proposition 2.3.2 and Remarks 2.3.3 and 2.3.4, we have the existence of a sequence of stopping times $\{\tau_j\}$ with $\tau_j \geq n$, $\tau = \lim_j \tau_j$, $j = 0, 1, \ldots$, and,

$$Z_n = \lim_j \uparrow \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_j}, Y_{\tau_j})|\hat{\mathcal{F}}_n]. \tag{2.3.5}$$

Then by the monotone convergence theorem for random variables[4], we have that,

$$\mathbb{E}_{\overline{\mathbb{P}}}[Z_n|\hat{\mathcal{F}}_{n-1}] = \lim_j \uparrow \mathbb{E}_{\overline{\mathbb{P}}}[\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_j}, Y_{\tau_j})|\hat{\mathcal{F}}_n]|\hat{\mathcal{F}}_{n-1}] = \lim_j \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_j}, Y_{\tau_j})|\hat{\mathcal{F}}_{n-1}] \leq Z_{n-1}. \tag{2.3.6}$$

Then from (2.3.2) and (2.3.6), we have

$$Z_n \geq \max\{F(\Pi_n, Y_n), \mathbb{E}_{\overline{\mathbb{P}}}[Z_{n+1}|\hat{\mathcal{F}}_n]\}, \tag{2.3.7}$$

which completes the forward proof.

Backward Proof.

Recall from (2.3.3) and the first equality in (2.3.4) that for any stopping time $\tau \geq n$, we have

$$F(\Pi_\tau, Y_\tau) = F(\Pi_n, Y_n) \cdot \mathbf{1}_{\{\tau=n\}} + F(\Pi_{\tau\vee(n+1)}, Y_{\tau\vee(n+1)}) \cdot \ \mathbf{1}_{(\tau>n)},$$

and,

$$\begin{aligned}\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)|\hat{\mathcal{F}}_n] &= \mathbf{1}_{\{\tau=n\}}\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_n, Y_n)|\hat{\mathcal{F}}_n] + \mathbf{1}_{(\tau>n)}\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau\vee(n+1)}, Y_{\tau\vee(n+1)})|\hat{\mathcal{F}}_n] \\ &= \mathbf{1}_{\{\tau=n\}}F(\Pi_n, Y_n) + \mathbf{1}_{(\tau>n)}\mathbb{E}_{\overline{\mathbb{P}}}[\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau\vee(n+1)}, Y_{\tau\vee(n+1)})|\hat{\mathcal{F}}_{n+1}]|\hat{\mathcal{F}}_n] \\ &\leq \mathbf{1}_{\{\tau=n\}}F(\Pi_n, Y_n) + \mathbf{1}_{(\tau>n)}\mathbb{E}_{\overline{\mathbb{P}}}[Z_{n+1}|\hat{\mathcal{F}}_n] \quad \text{(using(2.3.6) and (2.3.7))} \\ &= \max\{F(\Pi_n, Y_n), \ \mathbb{E}_{\overline{\mathbb{P}}}[Z_{n+1}|\hat{\mathcal{F}}_n]\}. \end{aligned} \tag{2.3.8}$$

If we take the supremum on the left hand side of (2.3.8), we obtain

$$Z_n \leq \max\{F(\Pi_n, Y_n), \mathbb{E}_{\overline{\mathbb{P}}}[Z_{n+1}|\hat{\mathcal{F}}_n]\}, \tag{2.3.9}$$

which ends the backward proof.

Thus from (2.3.7) and (2.3.9), we have that,

$$Z_n = \max\{F(\Pi_n, Y_n), \mathbb{E}_{\overline{\mathbb{P}}}[Z_{n+1}|\hat{\mathcal{F}}_n]\},$$

which ends the first part of the proof. We now proceed to the second part of the proof.

Using the monotone convergence theorem for random variables, we have from (2.3.5) that,

$$\mathbb{E}_{\overline{\mathbb{P}}}[Z_n] = \lim_j \uparrow \mathbb{E}_{\overline{\mathbb{P}}}[\mathbb{E}_{\overline{\mathbb{P}}}[(F(\Pi_{\tau_j}, Y_{\tau_j}))|\hat{\mathcal{F}}_n]] = \lim_j \uparrow \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau_j}, Y_{\tau_j})].$$

[4] The monotone convergence theorem for random variables states that; Let $0 \leq X_1 \leq X_2 \leq \ldots$ be a monotone non-decreasing sequence of non-negative random variables. Then $\mathbb{E}[\lim_{n\to\infty} X_n] = \lim_{n\to\infty} \mathbb{E}[X_n]$.

From which we have,

$$\mathbb{E}_{\overline{\mathbb{P}}}[Z_n] \leq \sup_{n\leq\tau\leq T} \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)] = V_n \tag{2.3.10}$$

But by the definition of snell envelope, we have for every $\tau > n$ that,

$$\mathbb{E}_{\overline{\mathbb{P}}}[Z_n] \geq \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)]. \tag{2.3.11}$$

Thus,

$$\mathbb{E}_{\overline{\mathbb{P}}}[Z_n] = \sup_{n\leq\tau\leq T} \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)] = V_n.$$

Lastly, we show that Z is the smallest supermartingale that dominates $F(\Pi_n, Y_n)$.

Suppose there exists another supermartingale Z' that dominates $F(\Pi_n, Y_n)$ such that $Z' \neq Z$, then using the optimal sampling theorem, we have that for every $\tau \geq n$, it follows that,

$$Z_n{}' \geq \mathbb{E}_{\overline{\mathbb{P}}}[Z_\tau{}'|\hat{\mathcal{F}}_n] \geq \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)|\hat{\mathcal{F}}_n] = Z_n.$$

This implies that $Z_n{}' \geq Z_n$.

Thus, $Z = \{Z_n\}$ is the smallest supermartingale that dominates $F(\Pi_n, Y_n)$. □

2.3.8 Theorem. *The supremum $V_0 = \sup_\tau \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_\tau, Y_\tau)]$ is attained if and only if the stopping time*

$$\tau^* = \inf\{n \geq 0 : Z_n = F(\Pi_n, Y_n)\},$$

is finite with probability one. When this happens, we say that τ^* is optimal.

Proof. Let us suppose that τ^* is finite almost surely. Then we make the assertion that the process $\{Z_{\tau^*\wedge n} : n \geq 0\}$ is a martingale with respect to the filtration $\{\hat{\mathcal{F}}_n\}$. The proof of the assertion is as follows

$$\begin{aligned}
\mathbb{E}_{\overline{\mathbb{P}}}[Z_{\tau^*\wedge(n+1)}|\hat{\mathcal{F}}_n] &= \mathbf{1}_{\{\tau^*\leq n\}} Z_{\tau^*} + \mathbf{1}_{\{\tau^*>n\}} \mathbb{E}_{\overline{\mathbb{P}}}[Z_{n+1}|\hat{\mathcal{F}}_n] \\
&= \mathbf{1}_{\{\tau^*\leq n\}} Z_{\tau^*} + \mathbf{1}_{\{\tau^*>n\}} Z_n \\
&= Z_{\tau^*\wedge n}.
\end{aligned}$$

Using Theorem 2.3.7 and the dominated convergence theorem[5,6], we have that,

$$V_0 = \mathbb{E}_{\overline{\mathbb{P}}}[Z_0] = \lim_n \mathbb{E}_{\overline{\mathbb{P}}}[Z_{\tau^*\wedge n}] = \mathbb{E}_{\overline{\mathbb{P}}}[Z_{\tau^*}] = \mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\tau^*}, Y_{t^*})].$$

Now suppose there is an optimal stopping stopping time $\overline{\tau}^*$, then our aim is to show that τ^* is finite. To show this, it is sufficient to show that $Z_{\overline{\tau}^*} = F(\Pi_{\overline{\tau}^*}, Y_{\overline{\tau}^*})$. But from Theorem 2.3.7, we have that, $F(\Pi_{\overline{\tau}^*}, Y_{\overline{\tau}^*}) \leq Z_{\overline{\tau}^*}$, and from the optional sampling theorem, we have that,

$$\mathbb{E}_{\overline{\mathbb{P}}}[F(\Pi_{\overline{\tau}^*}, Y_{\overline{\tau}^*})] = V_0 = \mathbb{E}[Z_0] \geq \mathbb{E}_{\overline{\mathbb{P}}}[Z_{\overline{\tau}^*}].$$

Thus we have the equality

$$F(\Pi_{\overline{\tau}^*}, Y_{\overline{\tau}^*}) = Z_{\overline{\tau}^*} \quad \text{almost surely.}$$

□

[5]The dominated convergence theorem states that: Let X_n be a random variable with $X_n \to X$ almost surely. Suppose there exists an integrable random variable Y such that $|X_n| \leq Y$ almost surely for all n, then $\lim_{n\to\infty} \mathbb{E}[X_n|G] = \mathbb{E}[X|G]$ almost surely, where G is a sigma algebra.

[6]We have also used the fact that for a martingale (X_i) and stopping time T with respect to a filtration $(\mathcal{F}_i)$, we have that, $\mathbb{E}[X_T] = \mathbb{E}[X_0]$ provided that $Pr(T < \infty) = 1$ and $\mathbb{E}[|X_T|] < \infty$.

The following result gives the dynamic programming equation for Theorem 2.3.7.

2.3.9 Proposition. Let $\mathcal{H}$ denote the Hilbert space. Then, the value function, $V(t, \Pi, y)$ is the unique viscosity solution of the following Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{cases} \max\{V_t + \mathcal{L}V - rV, F(t, \Pi, y) - V(t, \Pi, y)\} = 0 \ \text{ in } \mathcal{H} \\ V(t, \Pi, y) = F(t, \Pi, y) \ \text{ in } \mathcal{H}, \end{cases}. \tag{2.3.12}$$

and is bounded and locally lipschitz with respect to the Hilbert norm. Also, $V(t, \Pi, y)$ generally satisfies

$$\mathcal{L}V = \frac{1}{2}\text{tr}((D_t {D_t}^T + B_t {B_t}^T) D^2 V) + \langle C_t X_t, DV \rangle + A_t X_t \partial y \, V + B_t DV \cdot \partial y V. \tag{2.3.13}$$

Where D is the Fréchet derivative and $< . \, , \, . >$ represents the inner product in $\mathcal{H}$.

Using the approach from (Gatarek and Świech, 1999), it can be shown that the comparison argument holds for viscosity sub and super solutions of 2.3.9. It also holds in our case since by definition, the value function, V, is lower semicontinuous.

3. Application and Numerical Results

3.1 American Option with Stochastic Convenience Yield

In this section, we shall solve the optimal stopping problem for American option with stochastic convenience yield similar to the one considered in Ludkovski (2009). Consider the model given below.

$$\begin{cases} dX_t = -\kappa X_t dt + \sigma_X \left(\rho \, dW_t + \sqrt{1-\rho^2} \, dU_t\right) \\ dY_t = X_t dt + \sigma_Y dW_t, \end{cases} \tag{3.1.1}$$

where U_t and W_t are independent one dimensional Brownian motions.

Under the equivalent measure $\overline{\mathbb{P}}$, the linear diffusion process Y would be a martingale provided that we don't have a drift term. So that Y_t then satisfies the condition

$$Y_t = \mathbb{E}[X_t | \mathcal{F}_t^Y],$$

where $\mathbb{E}[\ \cdot\ |\mathcal{F}_t^Y]$ is the future conditional expectation. So under the measure $\overline{\mathbb{P}}$, we can convert Y_t into a martingale as follows

$$\begin{aligned} dY_t &= \sigma_Y \left[\frac{X_t}{\sigma_Y} \, dt + dW_t\right] \\ &= \sigma_Y \; d\left[\int_0^t \frac{X_t}{\sigma_Y} \, ds + W_t\right] \\ &= \sigma_Y \; d\overline{W}_t. \end{aligned} \tag{3.1.2}$$

Now, using (3.1.1) and (3.1.2), we can re-write our model in the form

$$\begin{cases} dX_t = -kX_t dt + \rho\sigma_X \, \dfrac{[dY_t - X_t dt]}{\sigma_Y} + \sigma_X \sqrt{1-\rho^2} \, dW^{\perp} \\ dY_t = \sigma_Y \; d\overline{W}_t, \end{cases} \tag{3.1.3}$$

where $\overline{W}_t$ and $W^{\perp}$ are independent wiener processes.

We would be dealing with the optimal stopping oproblem of the form

$$V_t = \sup_{t \le \tau \le T} \mathbb{E}[e^{-r\tau} f(X_\tau, Y_\tau)] = \sup_{t \le \tau \le T} \mathbb{E}[e^{-r\tau}(Y_\tau(c_1 + X_\tau) - c_2)_+], \quad c_i \in \mathbb{R}, \; i = 1, 2. \tag{3.1.4}$$

We can use the Kalman-Bucy filter described in chapter 2 to analyse the model in (3.1.1). Recall that in chapter 2, we obtained the filtering distribution Π_t in terms of the conditional variance E_t and the linear estimate $\hat{X}_t$ whose derivatives are given respectively as

$$dE_t = \left[(B_t \gamma_2 {B_t}^T + D_t \gamma_1 {D_t}^T) - (E_t {A_t}^T {H_t}^T (\Psi_t^2)^{-1} H_t \; A_t {E_t}^T) + C_t E_t + E_t {C_t}^T\right] dt \tag{3.1.5}$$

$$d\hat{X}_t = C_t \hat{X}_t dt + E_t {A_t}^T {H_t}^T (\Psi_t^2)^{-1} H_t [dY_t - A_t \hat{X}_t dt]. \tag{3.1.6}$$

We recall that, $H_t = (B_t {B_t}^T)^{-\frac{1}{2}}, \quad \gamma_1 = \sigma_1 {\sigma_1}^T$, and $\gamma_2 = \sigma_2 {\sigma_2}^T$.

Now comparing (3.1.1) with our model dynamics (2.1.1) and (2.1.2), we have that,

$$C_t = -\kappa,\ B_t = \rho\sigma_X = \sigma_Y,\ A_t = 1 \ \text{ and } \ D_t = \sigma_X\sqrt{1-\rho^2}.$$

So that in one dimension, we can re-write (3.1.6) as

$$d\hat{X}_t = -\kappa\hat{X}_t dt + \frac{E_t}{\sigma_Y\ t}\, d\overline{W}_t, \tag{3.1.7}$$

where,

$$d\overline{W}_t = \frac{dY_t - A_t\hat{X}_t dt}{\sigma_Y},$$

and the Brownian motion $d\overline{W}_t$ is the innovation process. Also, the variance, E_t can be determined and is a solution to the Ricatti ODE given as

$$dE_t = \left[-2\kappa E_t + \sigma_x^2 - \frac{{E_t}^2}{{\sigma_Y}^2\ t}\right] dt. \tag{3.1.8}$$

Under the measure $\overline{\mathbb{P}}$, we have that $\mathcal{F}_t^Y = \mathcal{F}_t^{\overline{W},\Pi_0}$, and

$$dY_t = \hat{X}_t dt + \sigma_Y d\overline{W}_t. \tag{3.1.9}$$

Thus, the filtered estimates; $\hat{X}_t$ and E_t, provide enough information for obtaining the conditional distribution $X_t|\mathcal{F}_t^Y$. The payoff for the exotic call option is given as

$$\begin{aligned}
\mathbb{E}[f(X_t,Y_t)|\mathcal{F}_t^Y] &= \mathbb{E}[(y(c_1+\hat{X}_t+\sqrt{E_t}\mathcal{X})-c_2)_+] \ \text{ where } \mathcal{X}\sim\mathcal{N}(0,1)\\
&= \int_{x^*}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}y\sqrt{E_t}\ x dx + ((c_1+\hat{X}_t)y - c_2)\int_{x^*}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\ dx\\
&= \frac{y\sqrt{E_t}}{\sqrt{2\pi}}e^{-\frac{(x^*)^2}{2}} + ((c_1+\hat{X}_t)y-c_2)(1-\Phi(x^*)) = F(\hat{X}_t,E_t,Y_t).
\end{aligned}$$

Where $\Phi(x^*)$ is the cumulative distribution function of the standard normal distribution with respect to x^* and,

$$x^* = \frac{c_2-(c_1+\hat{X}_t)y}{y\sqrt{E_t}}.$$

We can now reduce the problem in (3.1.4) to the problem

$$V(t,\hat{X},E,y) = \sup_{t\le\tau\le T}\mathbb{E}[e^{-r\tau}F(\hat{X}_\tau,E_\tau,Y_\tau)|\hat{X}_0 = x, E_0 = e, Y_0 = y]. \tag{3.1.10}$$

In order to solve the problem in (3.1.10), we shall make use of Proposition 2.3.9.

We first obtain (2.3.13) for the case of our problem. Using (3.1.7) and (3.1.9), we have the following:

$$\begin{aligned}
&B_t = 0.\\
&D_t{D_t}^T = \begin{bmatrix}\frac{E_t}{\sigma_Y t}\\ \\ \sigma_Y\end{bmatrix}\begin{bmatrix}\frac{E_t}{\sigma_Y t} & \sigma_Y t\end{bmatrix} = \begin{bmatrix}\frac{{E_t}^2}{{\sigma_Y}^2t^2} & \frac{E_t}{t}\\ \\ \frac{E_t}{t} & {\sigma_Y}^2\end{bmatrix}.
\end{aligned}$$

So that,

$$(D_tD_t^T)D^2V = \begin{bmatrix} \frac{E_t^2}{\sigma_Y{}^2t^2} & \frac{E_t}{t} \\ \frac{E_t}{t} & \sigma_Y{}^2 \end{bmatrix} \begin{bmatrix} V_{\hat{X}\hat{X}} & V_{\hat{X}y} \\ V_{\hat{X}y} & V_{yy} \end{bmatrix} = \begin{bmatrix} \frac{E_t^2}{\sigma_Y{}^2\, t^2}\, V_{\hat{X}\hat{X}} + \frac{E_t}{t}\, V_{\hat{X}y} & \frac{E_t^2}{\sigma_Y{}^2t^2}\, V_{\hat{X}y} + \frac{E_t}{t}\, V_{yy} \\ \frac{E_t}{t}\, V_{\hat{X}\hat{X}} + \sigma_Y{}^2\, V_{\hat{X}y} & \frac{E_t}{t} V_{\hat{X}y} + \sigma_Y{}^2\, V_{yy} \end{bmatrix}.$$

Also, we have

$$\langle c_tX_t, DV\rangle = \left\langle -\kappa\hat{X}_t, DV\right\rangle = [-\kappa\hat{X}, 0]\begin{bmatrix} V_{\hat{X}} \\ V_y \end{bmatrix} = [-\kappa\hat{X}V_{\hat{X}}] = -\kappa\hat{X}\ V_{\hat{X}}, \text{ and}$$

$$A_tX_t\ \partial y\ V = \hat{X}\ \partial y\ V = \hat{X}\ V_y.$$

Therefore, we can write

$$\begin{aligned} \mathcal{L}V &= \frac{1}{2}\left(\frac{E_t^2}{\sigma_Y{}^2\ t^2}\ V_{\hat{X}\hat{X}} + \frac{E_t}{t}\ V_{\hat{X}y} + \frac{E_t}{t}\ V_{\hat{X}y} + \sigma_Y{}^2\ V_{yy}\right) - \kappa\hat{X}V_{\hat{X}} + \hat{X}\ V_y \\ &= \frac{1}{2}\left(\frac{E_t^2}{\sigma_Y{}^2\ t^2}\ V_{\hat{X}\hat{X}}\right) + \frac{E_t}{t}\ V_{\hat{X}y} + \frac{1}{2}\sigma_Y{}^2\ V_{yy} - \kappa\hat{X}V_{\hat{X}} + \hat{X}\ V_y. \end{aligned}$$

Thus, using 2.3.9, we have

$$\begin{cases} \max\{V_t + \hat{X}V_y + \frac{1}{2}\sigma_Y{}^2V_{yy} - \kappa\hat{X}V_{\hat{X}} + \frac{1}{2}\frac{E_t^2}{\sigma_Y{}^2t^2}V_{\hat{X}\hat{X}} + \frac{E_t}{t}V_{\hat{X}y} - rV,\ F(\hat{X}, E, y) - V(t, \hat{X}, E, y)\} = 0 \\ V(T, \hat{X}, E, y) = F(\hat{X}, E, y). \end{cases} \tag{3.1.11}$$

3.2 Numerical Results

In order to solve (3.1.11), we first perform explicit finite difference discretization in space as follows

$$\begin{aligned} &\frac{dV}{dt} + \left(\frac{\hat{X}}{2\Delta y} + \frac{\sigma_Y{}^2}{2(\Delta y)^2}\right)V_{i,j+1} - \left(\frac{\hat{X}}{2\Delta y} - \frac{\sigma_Y{}^2}{2(\Delta y)^2}\right)V_{i,j-1} - \left(\frac{\sigma_Y{}^2}{(\Delta y)^2} + \frac{E_t^2}{\sigma_Y{}^2(\Delta t)^2(\Delta\hat{X})^2 + r}\right)V_{i,j} \\ &- \left(\frac{\kappa\hat{X}}{2\Delta\hat{X}} - \frac{E_t^2}{2\sigma_Y{}^2(\Delta t)^2(\Delta\hat{X})^2}\right)V_{i+1,j} + \left(\frac{\kappa\hat{X}}{2\Delta\hat{X}} + \frac{E_t^2}{2\sigma_Y{}^2(\Delta t)^2(\Delta\hat{X})^2}\right)V_{i-1,j} + \left(\frac{E_t}{4\Delta\hat{X}\Delta y\Delta t}\right)V_{i+1,j+1} \\ &- \left(\frac{E_t}{4\Delta\hat{X}\Delta y\Delta t}\right)V_{i+1,j-1} - \left(\frac{E_t}{4\Delta\hat{X}\Delta y\Delta t}\right)V_{i-1,j+1} + \left(\frac{E_t}{4\Delta\hat{X}\Delta y\Delta t}\right)V_{i-1,j+1} = 0, \end{aligned}$$

which can be represented in the form

$$\begin{aligned} &\frac{dV}{dt} + a_{i,j}V_{i,j+1} - b_{i,j}V_{i,j-1} - C_{i,j}V_{i,j} - d_{i,j}V_{i+1,j} + e_{i,j}V_{i-1,j} \\ &+ f_{i,j}V_{i+1,j+1} - g_{i,j}V_{i+1,j-1} - h_{i,j}V_{i-1,j+1} + l_{i,j}V_{i-1,j+1} = 0. \end{aligned}$$

So that,

$$\frac{dV_h}{dt} + A_hV_h = 0.$$

Where A_h is a matrix consisting of the coefficients of $V_{i,j}$ for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, M$. N and M are the number of space grid points in $\hat{X}$ and Y respectively.

Using the Euler-theta method, this reduces to

$$V_h{}^{n+1} = (I + \Delta t\ \theta A_h)^{-1}(I - \Delta t(1-\theta)\ A_h)V_h{}^n, \quad n = 0, 1, \ldots, N, \quad \text{with} \quad \theta \in [0, 1]. \tag{3.2.1}$$

The aim of the discretizations above is so as to obtain V from (3.2.1) and to reduce (3.1.11) to

$$\begin{cases} \max\{V(t, \hat{X}, E, y),\ F(\hat{X}, E, y) - V(t, \hat{X}, E, y)\} = 0 \\ V(T, \hat{X}, E, y) = F(\hat{X}, E, y). \end{cases} \tag{3.2.2}$$

We now proceed to evaluate (3.2.2) numerically and compare the graph of the intrinsic value F to that of the value function V.

To achieve our objective, we ran a code in Matlab with a 31×31 grid with 10 time intervals. We chose $E_0 = 0.0001,\ \hat{X}_0 = 0,$ and $Y_0 = 0$ and our time step size was taken to be $dt = T/M$. Below is a table showing the parameter values we used for our computation[7]

Parameter	c_1	c_2	r	$\sigma_{\hat{X}}$	σ_Y	κ	T
Value	1	2	0.1	0.3	0.1	2	1

Figure 3.1: Table of parameter values.

After our computation using Matlab, we obtained the following graphs:

[7] T = time to maturity, $\sigma_{\hat{X}}$ = volatility for the process $\hat{X}$, σ_Y = volatility for the process Y and r = stochastic convenience yield rate.

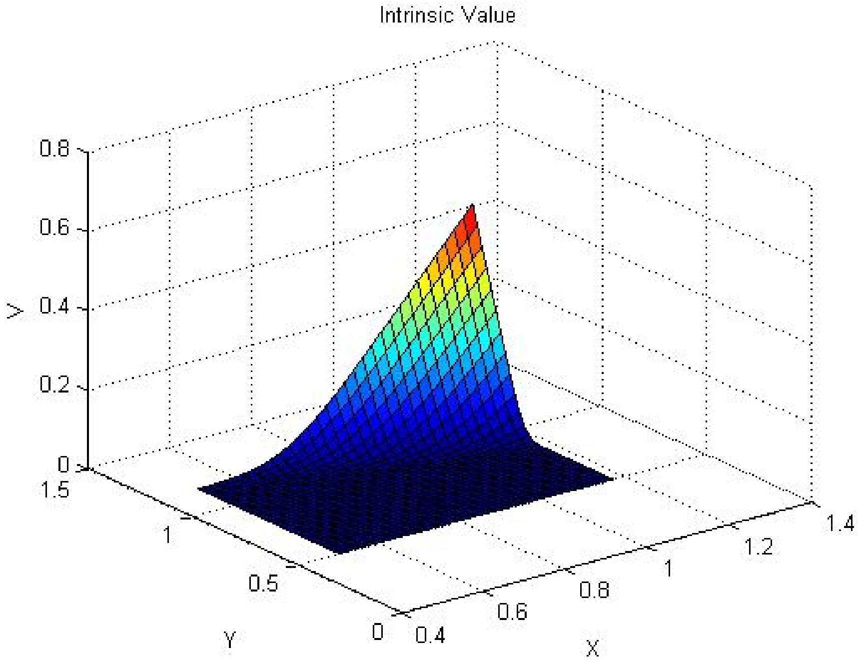


Figure 3.2: Plot of Intrinsic Value F for $\hat{X}_0 = 0$, $Y_0 = 0$, $E_0 = 0.0001$, $T = 1$, $r = 0.1$, $\kappa = 2$, $\sigma_{\hat{X}} = 0.3$, $\sigma_Y = 0.1$, $c1 = 1$, $c2 = 2$, $\theta = 1$, $N_X = 30$, $N_Y = 30$ and $M = 10$.

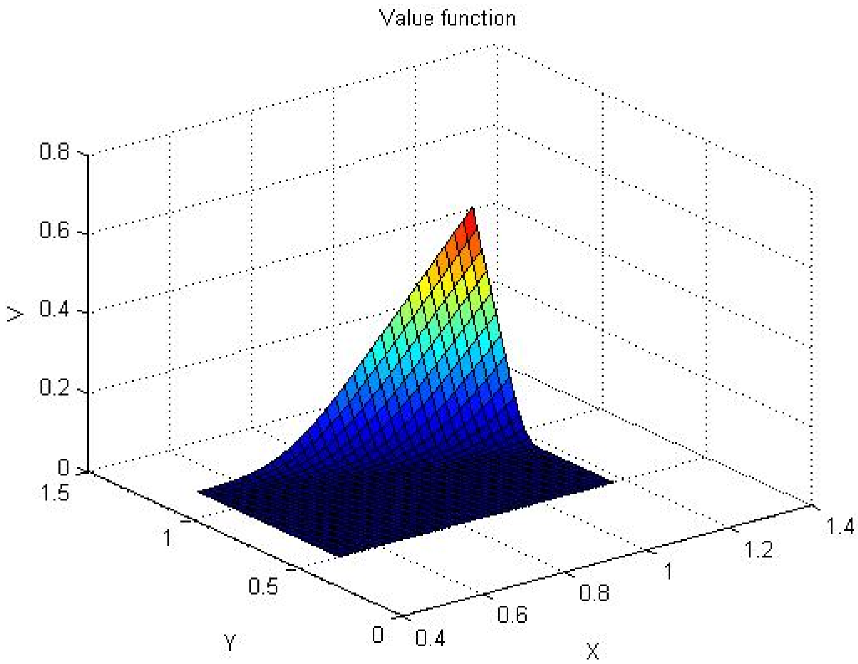


Figure 3.3: Plot of Value function V for $\hat{X}_0 = 0$, $Y_0 = 0$, $E_0 = 0.0001$, $T = 1$, $r = 0.1$, $\kappa = 2$, $\sigma_{\hat{X}} = 0.3$, $\sigma_Y = 0.1$, $c1 = 1$, $c2 = 2$, $\theta = 1$, $N_X = 30$, $N_Y = 30$ and $M = 10$.

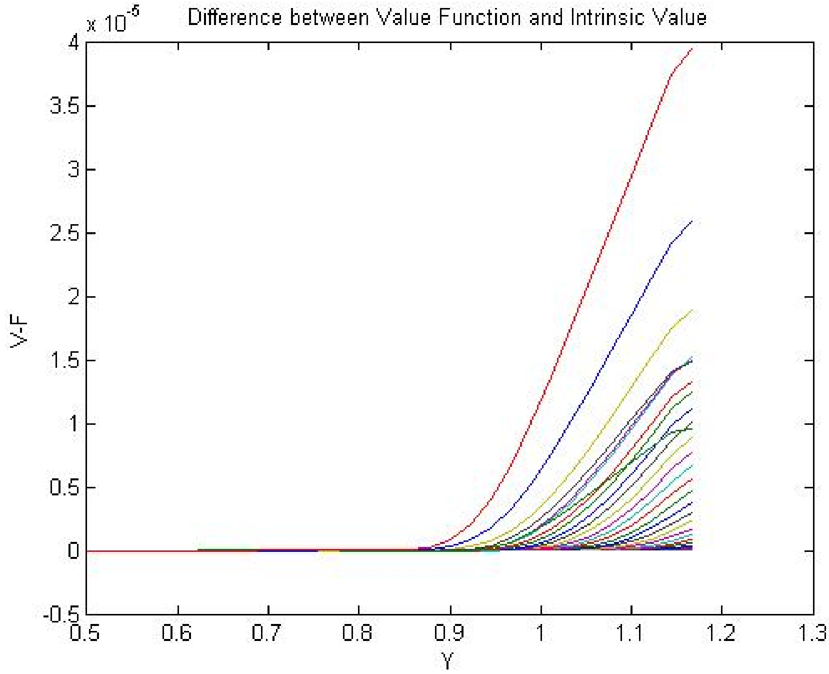


Figure 3.4: Plot of $V - F$

Figure 3.4 is a plot of the difference between the value function, V, and the intrinsic value, F. From this plot, we see that the value function V is indeed a concave envelope of the intrinsic value.

4. Conclusion

In this project work, we considered the optimal stopping problem of square-integrable Lévy processes under partial information. Using the Kalman-Bucy filter in Theorems 2.2.4 and 2.2.5, we converted this problem to an optimal stopping problem under full information by deriving a filtering distribution Π_t in the form of a finite linear estimate of the mean $\hat{X}_t$, and the variance E_t of the unobservable process $(X_t)_{0\leq t\leq T}$. We proposed a solution to the resulting optimal stopping problem under full information by using the Snell envelope theory and dynamic programming principle of Section 2.3. We further illustrated an application of this type of problem in mathematical finance. Using the Kalman-bucy formulation, we analysed the model considered in the application and using Proposition 2.3.9, we obtained the HJB equation for which the value function is a viscosity solution. Then using the explicit finite difference discretization together with the Euler-theta method, we obtained our value function. Finally, we implemented (3.1.11) in Matlab and plotted the value function, the intrinsic value and their difference on separate plots. From the plots, we observed that the value function is a concave envelope of the payoff of the American option with stochastic convenience yield.

Acknowledgements

My profound gratitude goes to the Alpha and the Omega, my source of insight and aplomb. To my supervisors, Dr. Denis Ikpe and Mr. Edward Aidoo, I express my sincere gratitude for your guidance, patience and accessibility. Working with you was a good experience.

To my mom and siblings, I'm really appreciative of your love and support and to my colleagues, thanks for being a part of my story. I'm certain we'll meet at the acme. SHALOM.

References

D. Applebaum and S. Blackwood. The kalman-bucy filter for integrable lévy processes with infinite second moment. *Journal of Applied Probability*, 52(03):636–648, 2015.

A. Bensoussan. *Stochastic control of partially observable systems*. Cambridge University Press, 2004.

S. Blackwood. *Lévy processes and filtering theory*. PhD thesis, University of Sheffield, 2014.

D. Crisan. Particle approximations for a class of stochastic partial differential equations. *Applied Mathematics & Optimization*, 54(3):293–314, 2006.

P. Fitzsimmons, J. Pitman, and M. Yor. Markovian bridges: construction, palm interpretation, and splicing. In *Seminar on Stochastic Processes, 1992*, pages 101–134. Springer, 1993.

D. Gatarek and Świech. Optimal stopping in hilbert spaces and pricing of american options. *Mathematical methods of operations research*, 50(1):135–147, 1999.

T. P. Hill. Knowing when to stop: how to gamble if you must—the mathematics of optimal stopping. *American Scientist*, 97(2):126–133, 2009.

D. Ikpe and R. Becker. Kalman filtering in the information-based credit risky asset pricing framework. Preprint, 2016.

U. Jensen and G. H. Hsu. Optimal stopping by means of point process observations with applications in reliability. *Mathematics of Operations Research*, 18(3):645–657, 1993.

I. Karatzas and S. E. Shreve. *Methods of mathematical finance*, volume 1. Springer, 1998.

M. Ludkovski. A simulation approach to optimal stopping under partial information. *Stochastic Processes and their Applications*, 119(12):4061–4087, 2009.

A. Papapantoleon. An introduction to lévy processes with applications in finance. *arXiv preprint arXiv: 0804.0482*, 2008.

S. Peszat and J. Zabczyk. *Stochastic partial differential equations with Lévy noise: An evolution equation approach*, volume 113. Cambridge University Press, 2007.

B. Rémillard and J.-F. Renaud. A martingale representation for the maximum of a lévy process. *Commun Stoch Anal*, 5(4):683–688, 2011.

R. Rishel and K. Helmes. A variational inequality sufficient condition for optimal stopping with application to an optimal stock selling problem. *SIAM journal on control and optimization*, 45(2):580–598, 2006.

M. Vannestål. *Optimal stopping and the American put under incomplete information*. PhD thesis, Master Thesis. Uppsala University, 2011.

G. Wang and Z. Wu. Kalman–bucy filtering equations of forward and backward stochastic systems and applications to recursive optimal control problems. *Journal of Mathematical Analysis and Applications*, 342(2):1280–1296, 2008.

Z. Wang. *Optimal stopping and switching problems with financial applications*. PhD thesis, Columbia University, 2017.

P. Wieslander. On monte carlo approximation of the snell envelope with application to the pricing of american options. 2014.

F. Ye and E. Zhou. Optimal stopping of partially observable markov processes: a filtering-based duality approach. *IEEE Transactions on Automatic Control*, 58(10):2698–2704, 2013.

E. Zhou. Optimal stopping under partial observation: Near-value iteration. *IEEE Transactions on Automatic Control*, 58(2):500–506, 2013.